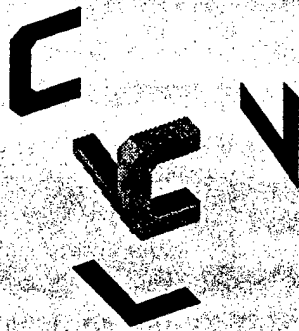# Understanding Object Motion

Zoran Duric[1,2], Ehud Rivlin[1,3], Azriel Rosenfeld[1]

[1]Center for Automation Research
University of Maryland
College Park, MD 20742-3275

[2]Machine Learning and Inference Laboratory
George Mason University
Fairfax, VA 22030-4444

[3]Department of Computer Science
Technion — Israel Institute of Technology
Haifa, Israel 32000

COMPUTER VISION LABORATORY

CENTER FOR AUTOMATION RESEARCH

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742-3275

Understanding Object Motion

Zoran Duric[1,2], Ehud Rivlin[1,3], Azriel Rosenfeld[1]

[1]Center for Automation Research
University of Maryland
College Park, MD 20742-3275

[2]Machine Learning and Inference Laboratory
George Mason University
Fairfax, VA 22030-4444

[3]Department of Computer Science
Technion – Israel Institute of Technology
Haifa, Israel 32000

Abstract

Many types of common objects, such as tools and vehicles, usually move in simple ways when they are wielded or driven: The natural axes of the object tend to remain aligned with the local trihedron defined by the object's trajectory. Based on this observation we use a model called *Frenet-Serret motion* which corresponds to the motion of a moving trihedron along a space curve. Knowing how the Frenet-Serret frame is changing relative to the observer gives us essential information for understanding the object's motion. This is illustrated here for four examples, involving tools (a wrench and a saw) and vehicles (an accelerating van, a turning taxi).

1 Introduction

An object moves because it is self-propelled (e.g., a vehicle) or because it is wielded (or thrown[1]) by an agent (e.g., a tool). Motion that efficiently performs a locomotional or mechanical function requires efficient energy transfer from the vehicle's engine or the agent's arm to the object, in order to efficiently overcome the constraints imposed by the environment in which the motion takes place (air resistance, friction, etc.). Assuming that an object has natural axes (e.g. the long axis of a stick), efficient force transfer requires simple relationships between the natural axes of the object and the motion trajectory. These relationships insure that the object can perform its function efficiently.

The most general model of object motion is unrestricted rigid motion. This type of motion is not common in everyday life. Usually objects are supported, and motion takes place when an object is in contact with a surface, another object, or an agent. In these cases (tool acting on a recipient object; ground vehicle) the motion becomes interestingly constrained.

In our work we consider the relationship between this constrained motion and the object's geometry. To analyze this relationship we use two frames: the object frame and the frame of the motion trajectory. "Efficient" motion calls for a simple relationship between the object frame and the motion frame, and this relationship remains constant during the motion. Based on this observation we use a model called *Frenet-Serret motion* which corresponds to the motion of a moving trihedron along a space curve [8]. The parameters of the motion are given by the curvature and torsion of the space curve along which the object moves.

In practice the simple nature of the environment in which the object moves provides further constraints. A ground vehicle is moving on relatively flat terrain, and a tool is often acting on a planar surface. The motion is mostly planar (though the plane might rotate slightly through the motion). Over a long time period the motion is Frenet-Serret and over a short time period the motion is approximately planar and often approximately translational.

We use the relationship between the object frame and the motion frame to analyze image sequences. Given a sequence of images of the moving object, our analysis enables us to output the motion and trajectory parameters of the object. Knowing how the Frenet-Serret frame is changing relative to the observer gives us essential information for understanding the object's motion. Our analysis can also handle constraints on the motion. For example, the parameters of the object's trajectory depend on its speed, mass, size, and on the medium through which it moves. These factors impose bounds on the curvature and torsion of the trajectory.

In this paper we approach object motion understanding through analysis of long image sequences. A key question in this context is how to relate short-sequence motion estimation to long-sequence motion estimation. Using the Frenet-Serret frame provides us with an ability to understand motion over a long time period. We can derive the motion parameters from the parameters of the trajectory and obtain motion descriptions suitable for long sequence analysis. Using these tools we can show, for example, that rotation becomes significant only in long

---

[1]We assume in this paper that the propulsive force is applied to the object continuously, unlike the case of a projectile where it is applied only initially. We will not discuss projectiles further here.

sequences, and that in a short sequence translation is usually dominant. We show that using simplified scene and imaging models we can get adequate local estimates (short sequence, 2-4 frames) by analyzing the images, and by observing these estimates over a long sequence we can accumulate them to describe the object's trajectory. Analysis of the trajectory parameters provides us with tools for understanding long-term object motion.

2 Related Work

Understanding object motion is based on extracting the object's motion parameters from an image sequence. Broida and Chellappa [1] proposed a framework for motion estimation of a vehicle using Kalman filtering. Weng et al. [16] assumed an object that possesses an axis of symmetry, and a constant angular momentum model which constrained the motion over a local frame subsequence to be a superposition of precession and translation. The trajectory of the center of rotation can be approximated by a vector polynomial. Changing the parameters of the model with time allows adaptation to long-term changes in the motion characteristics. Their work was based on correspondence; at least eight pairs of corresponding points were needed.

Accumulating the information obtained from the motion analysis of the sequence to achieve an estimate of the moving object's trajectory is another step toward understanding object motion. (A good survey of motion-based recognition was compiled by Cedras and Shah [5].) Bruckstein et al. [2, 3] assumed a known object model (a rigid rod or disk) and tried to recover the object's trajectory and rotation. They showed that five images are enough to recover the motion of a rod or a disk in accordance with physical laws. Techniques from algebraic geometry were used to establish the existence of solutions to the resulting polynomial equations.

Engel and Rubin [9] (and similarly Gould and Shah [11]) used motion characteristics obtained by tracking representative points on an object to identify important events corresponding to changes in direction, speed and acceleration in the object's motion.

Work has also been done on higher-level descriptions of object trajectories in terms of such concepts as stopping/starting, object interactions, and motion verbs[4, 12, 13]. This level of object motion description will not be treated in this paper.

In [6] Duric et al. tried to determine the function of an object from its motion. Given a sequence of images of a known object performing some function, they attempted to determine what that function was. They showed that the motion of an object, when combined with information about the object and its uses, provides strong constraints on the possible function being performed. Their flow-based analysis treated relatively short sequences.

In this paper a model for object trajectory analysis is used, and a constant relationship between the object frame and the motion frame is established. The use of the Frenet-Serret frame provides a vocabulary appropriate for describing longer motion sequences.

3 Motion Models

3.1 Rigid Body Motion

To facilitate the derivation of the motion equations of a rigid body $\mathcal{B}$ we use two rectangular coordinate frames, one $(Oxyz)$ fixed in space, the other $(Cx_1y_1z_1)$ fixed in the body and moving with it. The coordinates X_1, Y_1, Z_1 of any point P of the body with respect to the moving frame are constant with respect to time t, while the coordinates X, Y, Z of the same point P with respect to the fixed frame are functions of t. It is assumed that these functions are differentiable with respect to t. The position of the moving frame at any instant is given by the position $\vec{d_c} = (X_c \ Y_c \ Z_c)^T$ of the origin C, and by the nine direction cosines of the axes of the moving frame with respect to the fixed frame. Let $\vec{\imath}$, $\vec{\jmath}$, and $\vec{k}$ be the unit vectors in the directions of the Ox, Oy, and Oz axes, respectively; and let $\vec{\imath}_1$, $\vec{\jmath}_1$, and $\vec{k}_1$ be the unit vectors in the directions of the Cx_1, Cy_1, and Cz_1 axes, respectively. For a given position $\vec{p}$ of P in $Cx_1y_1z_1$ we have the position $\vec{r_p}$ of P in $Oxyz$:

$$\vec{r_p} \equiv \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \vec{\imath}\cdot\vec{\imath}_1 & \vec{\imath}\cdot\vec{\jmath}_1 & \vec{\imath}\cdot\vec{k}_1 \\ \vec{\jmath}\cdot\vec{\imath}_1 & \vec{\jmath}\cdot\vec{\jmath}_1 & \vec{\jmath}\cdot\vec{k}_1 \\ \vec{k}\cdot\vec{\imath}_1 & \vec{k}\cdot\vec{\jmath}_1 & \vec{k}\cdot\vec{k}_1 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} + \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} \equiv R\vec{p} + \vec{d_c} \tag{1}$$

where R is the matrix of the direction cosines (the frames are taken as right-handed so that $\det R = 1$). If we differentiate (1) with respect to time and use the fact that $\vec{p} = R^T(\vec{r_p} - \vec{d_c})$, we obtain

$$\dot{\vec{r_p}} = \dot{R}\vec{p} + \dot{\vec{d_c}} = \dot{R}R^T(\vec{r_p} - \vec{d_c}) + \dot{\vec{d_c}} \equiv \Omega(\vec{r_p} - \vec{d_c}) + \dot{\vec{d_c}}. \tag{2}$$

The skew matrix $\Omega = \dot{R}R^T = -R\dot{R}^T$ is the rotational velocity matrix and $\dot{\vec{d_c}}$ is the translational velocity vector. Multiplying a vector $(\vec{r_p} - \vec{d_c})$ by the skew matrix Ω can be replaced by taking the cross product $\vec{\omega} \times (\vec{r_p} - \vec{d_c})$ where $\vec{\omega} = (\omega_x \ \omega_y \ \omega_z)^T$ is the rotational velocity vector.

3.2 Motion along a Smooth Curve

Consider a moving frame $Cx_1y_1z_1$ (fixed in a rigid body $\mathcal{B}$), which moves with C along a space curve Γ while rotating so that the Cx_1 and Cy_1 axes concide with, respectively, the tangent and principal normal of Γ. This means that as C moves along Γ the $Cx_1y_1z_1$ frame concides with the Frenet-Serret trihedron at C: $Ctnb$. This trihedron consists of the tangent $\vec{t}$, the principal normal $\vec{n}$, and the binormal $\vec{b}$, which are mutually orthogonal (see Figure 1). The geometry of this motion is completely defined by Γ.

Let $\vec{d_\gamma}(s)$ denote the position of C, in the fixed coordinate frame $Oxyz$, when it has moved along Γ through a total arc length of s. For any position $\vec{p}$ of a point P on $\mathcal{B}$ in $Ctnb$, the position $\vec{r_p}$ in $Oxyz$ is given by (1) with the matrix of direction cosines R suitably defined (see Figure 1). If $\vec{t} = (t_1 \ t_2 \ t_3)^T$, $\vec{n} = (n_1 \ n_2 \ n_3)^T$ and $\vec{b} = (b_1 \ b_2 \ b_3)^T$ are the unit vectors along Ct, Cn and Cb, differential geometry gives us

$$\vec{t} = \vec{d}_\gamma', \quad \vec{n} = \kappa^{-1}\vec{d}_\gamma'', \quad \vec{b} = \vec{t} \times \vec{n}, \tag{3}$$
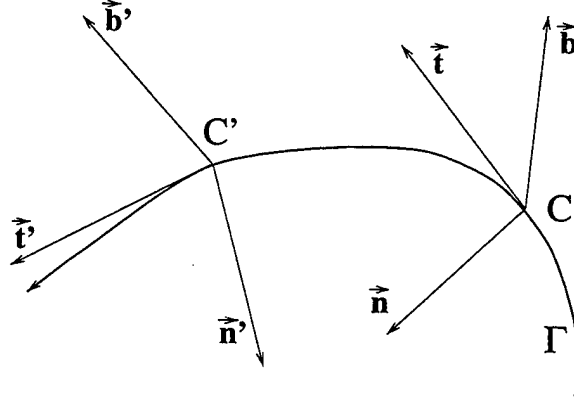
3

Figure 1: The Frenet-Serret coordinate frame moves along the path Γ.

where κ is the curvature of Γ. Then we have

$$R = \begin{pmatrix} t_1 & n_1 & b_1 \\ t_2 & n_2 & b_2 \\ t_3 & n_3 & b_3 \end{pmatrix}. \tag{4}$$

We have the Frenet-Serret formulas [14]

$$\vec{t}' = \kappa\vec{n}, \quad \vec{n}' = -\kappa\vec{t} + \tau\vec{b}, \quad \vec{b}' = -\tau\vec{n} \tag{5}$$

where τ is the torsion of Γ. Using (4) and (5), equation (2) can be written as

$$\vec{r}_p' = \vec{\omega}_d \times (\vec{r}_p - \vec{d}_\gamma) + \vec{t} \tag{6}$$

where the Darboux vector $\vec{\omega}_d = \tau\vec{t} + \kappa\vec{b}$ is the rotational velocity vector and the unit tangent $\vec{t}$ of Γ is the translational velocity vector; the motion parameter is the arc length s. If, instead of using arc length as a motion parameter, time t is used, the rotational velocity $\vec{\omega}_d$ and translational velocity $\vec{t}$ are scaled by the speed $v = ds/dt$ of the point C. In that case the equation of motion becomes

$$\dot{\vec{r}}_p = v\vec{\omega}_d \times (\vec{r}_p - \vec{d}_\gamma) + v\vec{t}. \tag{7}$$

In the special case where Γ is a plane curve we have $\tau = 0$ (Γ is torsionless), and thus $\vec{\omega}_d = \kappa\vec{b}$. Equation (7) then becomes

$$\dot{\vec{r}}_p = v\kappa\vec{b} \times (\vec{r}_p - \vec{d}_\gamma) + v\vec{t}. \tag{8}$$

3.3   Simple Motions of Objects

Objects move in reaction to forces which are being applied to them. When the forces acting on an object are added, the resultant force $\vec{F}$ determines the direction of motion and the moments of the forces (or the torques) determine the rotation of the object. If the force $\vec{F}$ is applied to the object $\mathcal{B}$ at the point P, the moment $\vec{M}$ is given by $\vec{M} = \vec{r}_p \times \vec{F}$ where $\vec{r}_p$ is the position of

4

P relative to a point C. $\vec{M}$ has the same direction as the axis of the rotation of $\mathcal{B}$ that results from applying $\vec{F}$.

The engine of a vehicle needs to apply force to the vehicle in order to move it from one position to another. If the path is prespecified (as in the case of a ground vehicle on a road), efficient application of the force requires that the angles between the instantaneous directions of the force and the directions of the path elements be small. The force differential generates torques which help turn the vehicle around the axis of rotation normal to the (osculating) plane of the path. During a turn, the wheels rotate with different speeds; the greater the distance between the wheels the larger their difference in speed. In order to minimize this difference the distance between the wheels needs to be small. Also, when forces are applied to the wheels the resulting torques are larger when the vehicle is moving along a short axis; but these torques need to be as small as possible to improve the handling of and minimize stresses on the vehicle. Because of all these factors the principal axis of inertia of the vehicle should be tangent to the path of the vehicle. It should be pointed out that [7] the translational velocity at any point on a ground vehicle is typically orders of magnitude larger than its rotational velocity (around the vehicle's center of mass). The rotational velocity becomes significant only when the vehicle is observed over a significant period of time (at least several frames).

In the case of a moving tool the force is used not only to move the tool, but to act on a recipient object. Therefore, the required force depends on the task. For example, sawing involves continuously exerting a force perpendicular to the path of the saw; tightening with a wrench involves continuously exerting torque around the axis of rotation. (Note that the force may not be applied to the recipient object continuously; for example, when we swing a hammer, the force is applied only when the head of the hammer hits the object.) Developing a general theory of tool motion is a subject of our continuing research.

4 Computing Motion from Image Sequences

For the purpose of estimating object motion from images we rewrite equation (2) in the following way:

$$\dot{\vec{r}}_p = \vec{\omega} \times (\vec{r}_p - \vec{d}_c) + \dot{\vec{d}}_c = \vec{\omega} \times \vec{r}_p + \vec{T} \qquad (9)$$

where $\vec{T} = \dot{\vec{d}}_c - \vec{\omega} \times \vec{d}_c \equiv (U \ V \ W)^T$ is the translational velocity expressed in the fixed (camera) coordinate frame $Oxyz$. We will later show how the translational velocity $\dot{\vec{d}}_c$ can be recovered from $\vec{T}$.

4.1 The Imaging Models

Let (X, Y, Z) denote the Cartesian coordinates of a scene point with respect to the fixed camera frame (see Figure 2), and let (x, y) denote the corresponding coordinates in the image plane. The equation of the image plane is $Z = f$, where f is the focal length of the camera. The

perspective projection onto this plane is given by

$$x = \frac{fX}{Z}, \quad y = \frac{fY}{Z}. \tag{10}$$

For weak perspective projection we need a reference point (X_c, Y_c, Z_c). A scene point (X, Y, Z) is first projected onto the point (X, Y, Z_c); then, through plane perspective projection, the point (X, Y, Z_c) is projected onto the image point (x, y). The projection equations are then given by

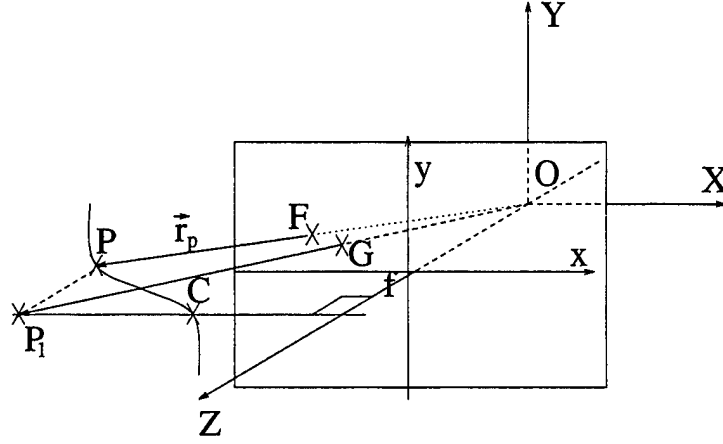$$x = \frac{X}{Z_c}f, \quad y = \frac{Y}{Z_c}f. \tag{11}$$

Figure 2: The plane perspective projection image of P is $F = f(X/Z, Y/Z, 1)$; the weak perspective projection image of P is obtained through the plane perspective projection of the intermediate point $P_1 = (X, Y, Z_c)$ and is given by $G = f(X/Z_c, Y/Z_c, 1)$.

4.2   The Image Motion Field and the Optical Flow Field

The instantaneous velocity of the image point (x, y) under perspective projection is obtained by taking the derivatives of (10) and using (9):

$$\dot{x} = \frac{\dot{X}Z - X\dot{Z}}{Z^2} = \frac{Uf - xW}{Z} - \omega_x \frac{xy}{f} + \omega_y \left(\frac{x^2}{f} + f\right) - \omega_z y, \tag{12}$$

$$\dot{y} = \frac{\dot{Y}Z - Y\dot{Z}}{Z^2} = \frac{Vf - yW}{Z} - \omega_x \left(\frac{y^2}{f} + f\right) + \omega_y \frac{xy}{f} + \omega_z x. \tag{13}$$

The instantaneous velocity of the image point (x, y) under weak perspective projection can be obtained by taking derivatives of (11) with respect to time and using (9):

$$\dot{x} = f\frac{\dot{X}Z_c - X\dot{Z}_c}{Z_c^2} = \frac{Uf - xW}{Z_c} + f\omega_y \frac{Z}{Z_c} - \omega_z y, \tag{14}$$

$$\dot{y} = f\frac{\dot{Y}Z_c - Y\dot{Z}_c}{Z_c^2} = \frac{Vf - yW}{Z_c} - f\omega_x \frac{Z}{Z_c} + \omega_z x. \tag{15}$$

6

Let $\vec{\imath}$ and $\vec{\jmath}$ be the unit vectors in the x and y directions, respectively; $\dot{\vec{r}} = \dot{x}\vec{\imath} + \dot{y}\vec{\jmath}$ is the projected motion field at the point $\vec{r} = x\vec{\imath} + y\vec{\jmath}$. If we choose a unit direction vector $\vec{n}_r$ at the image point $\vec{r}$ and call it the normal direction, then the *normal motion field* at $\vec{r}$ is $\dot{\vec{r}}_n = (\dot{\vec{r}} \cdot \vec{n}_r)\vec{n}_r$. $\vec{n}_r$ can be chosen in various ways; the usual choice (as we shall now see) is the direction of the image intensity gradient.

Let $I(x, y, t)$ be the image intensity function. The time derivative of I can be written as

$$\frac{dI}{dt} = \frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = (I_x\vec{\imath} + I_y\vec{\jmath}) \cdot (\dot{x}\vec{\imath} + \dot{y}\vec{\jmath}) + I_t = \nabla I \cdot \dot{\vec{r}} + I_t$$

where ∇I is the image gradient and the subscripts denote partial derivatives.

If we assume $dI/dt = 0$, i.e. that the image intensity does not vary with time, then we have $\nabla I \cdot \vec{u} + I_t = 0$. The vector field $\vec{u}$ in this expression is called the *optical flow*. If we choose the normal direction $\vec{n}_r$ to be the image gradient direction, i.e. $\vec{n}_r \equiv \nabla I / \|\nabla I\|$, we then have

$$\vec{u}_n = (\vec{u} \cdot \vec{n}_r)\vec{n}_r = \frac{-I_t \nabla I}{\|\nabla I\|^2} \tag{16}$$

where $\vec{u}_n$ is called the *normal flow*.

It was shown in [15] that the magnitude of the difference between $\vec{u}_n$ and the normal motion field $\dot{\vec{r}}_n$ is inversely proportional to the magnitude of the image gradient. Hence $\dot{\vec{r}}_n \approx \vec{u}_n$ when $\|\nabla I\|$ is large. Equation (16) thus provides an approximate relationship between the 3-D motion and the image derivatives. We will use this approximation later in this paper.

5 Tool Motion

We assume that the tool is (approximately) planar and that its velocity is composed of a translational velocity in the plane of the tool and a rotational velocity around an axis orthogonal to the plane of the tool.

5.1 The Image Motion Field of a Wielded Tool

Let the normal to the plane be $\vec{N} = (N_x \ N_y \ N_z)^T$; the equation of the plane orthogonal to $\vec{N}$ which passes through the point $(0, 0, Z_0)$ on the z-axis of the $Oxyz$ coordinate frame is given by

$$XN_x + YN_y + (Z - Z_0)N_z = 0. \tag{17}$$

If we assume a nondegenerate view (i.e., $N_z > 0$) for points on the plane we obtain from (17) and (10)

$$\frac{1}{Z} = \frac{1}{fZ_0}\left(f + f\frac{XN_x}{ZN_z} + f\frac{YN_y}{ZN_z}\right) = \frac{1}{fZ_0}(f + px + qy) \tag{18}$$

where $p = N_x N_z^{-1}$ and $q = N_y N_z^{-1}$. From our assumption about rotational velocity it follows that we have $\vec{\omega} = (p\omega_z \ q\omega_z \ \omega_z)$ for some ω_z. Also, since we have assumed that the translation is in the plane of the tool we have $\vec{N} \cdot \vec{T} = 0$, or equivalently

$$(p \ q \ 1)^T \cdot (U \ V \ W)^T = Up + Vq + W = 0.$$

It follows that we have

$$W = -Up - Vq. \tag{19}$$

From (12–13), (18), and (19) we obtain the equations of projected motion for points on the plane:

$$\dot{x} = \frac{Uf + xUp + xVq}{fZ_0}(f + px + qy) - p\omega_z \frac{xy}{f} + q\omega_z \left(\frac{x^2}{f} + f\right) - \omega_z y, \tag{20}$$

$$\dot{y} = \frac{Vf + yUp + yVq}{fZ_0}(f + px + qy) - p\omega_z \left(\frac{y^2}{f} + f\right) + q\omega_z \frac{xy}{f} + \omega_z x. \tag{21}$$

Equations (20-21) relate the image (projected) motion field to the scaled components of the translational velocity $Z_0^{-1}U \equiv U_0$ and $Z_0^{-1}V \equiv V_0$, the rotational parameter ω_z, and the normal to the plane $(p \ q \ 1)^T$.

Given the point $\vec{r} = x\vec{i} + y\vec{j}$ and the normal direction $n_x\vec{i} + n_y\vec{j}$, from (20-21) the normal motion field $\dot{\vec{r}}_n \cdot \vec{n} = n_x\dot{x} + n_y\dot{y}$ is given by

$$
\begin{aligned}
\dot{\vec{r}}_n \cdot \vec{n} &= U_0(f + px + qy)[n_x + (xn_x + yn_y)pf^{-1}] + V_0(f + px + qy)[n_y + (xn_x + yn_y)qf^{-1}] \\
&\quad + \omega_z[n_x(-y + qf - pxyf^{-1} + qx^2f^{-1}) + n_y(x - pf + qxyf^{-1} - py^2f^{-1})] \\
&\equiv U_0\varphi_1(p, q; \vec{r}, \vec{n}) + V_0\varphi_2(p, q; \vec{r}, \vec{n}) + \omega_z\varphi_3(p, q; \vec{r}, \vec{n})
\end{aligned} \tag{22}
$$

where the φs are nonlinear functions of p, q, $\vec{r}$, and $\vec{n}$ is given by

$$\varphi_1(p, q; \vec{r}, \vec{n}) = (f + px + qy)[n_x + (xn_x + yn_y)pf^{-1}], \tag{23}$$

$$\varphi_2(p, q; \vec{r}, \vec{n}) = (f + px + qy)[n_y + (xn_x + yn_y)qf^{-1}], \tag{24}$$

$$\varphi_3(p, q; \vec{r}, \vec{n}) = n_x(-y + qf - pxyf^{-1} + qx^2f^{-1}) + n_y(x - pf + qxyf^{-1} - py^2f^{-1}). \tag{25}$$

In Equation (22) $\vec{r}$ and $\vec{n}$ are observable from images, while the 5–tuple $(p, q, U_0, V_0, \omega_z)$ is not directly observable. To estimate this 5–tuple we need estimates of $\dot{\vec{r}}_n \cdot \vec{n}$ at five or more image points.

5.2 Estimating Tool Motion from Normal Flow

If we use the spatial image gradient as the normal direction $\vec{n}_r \equiv \nabla I/\|\nabla I\| = n_x\vec{i} + n_y\vec{j}$ and $\dot{\vec{r}}_n \approx \vec{u}_n$ we can obtain an approximate equation corresponding to (22) by replacing the left hand side of (22) by the normal flow $-I_t/\|\nabla I\|$. This equation involves the eight unknown elements

8

of **c**. For each point (x_i, y_i), $i = 1, \ldots, m$ of the image at which $\|\nabla I(x_i, y_i, t)\|$ is large we can write one such equation. If we have m such points, where $m \gg 5$, we have an over-determined system of equations

$$\Phi(p, q) \cdot (U_0 \ V_0 \ \omega_z)^T \approx \mathbf{b} \tag{26}$$

where the $m \times 3$ matrix function Φ is given by

$$\Phi(p, q) = [\varphi_1(p, q); \varphi_2(p, q); \varphi_3(p, q)]$$

(i.e., its columns are m-vectors that correspond to values of ϕ at points (x_i, y_i)), and the elements of the m-vector $\mathbf{b}$ are $-(\partial I(x_i, y_i, t)/\partial t)/\|\nabla I(x_i, y_i, t)\|$.

We seek the solution of the system (26) for which $\|\mathbf{b} - \Phi \cdot (U_0 \ V_0 \ \omega_z)^T\|$ is minimal — i.e., we are seeking the solution of (26) in the least squares sense. This is a separable nonlinear least squares problem; a good stable solution and an algorithm were given by Golub and Pereyra in [10]. It was shown that the problem is equivalent to minimizing

$$r(p, q) = \|\mathbf{b} - \Phi(p, q)\Phi^+(p, q)\mathbf{b}\|, \tag{27}$$

where Φ^+ is the generalized inverse of Φ. $r(p, q)$ is first minimized to obtain optimal values $\hat{p}$ and $\hat{q}$ of p and q respectively; these values are then used to obtain $\Phi(\hat{p}, \hat{q})$. The linear least squares method is then used to minimize $\|\mathbf{b} - \Phi(\hat{p}, \hat{q}) \cdot (U_0 \ V_0 \ \omega_z)^T\|$ and obtain optimal values of the motion parameters U_0, V_0, and ω_z. After estimating p, q, U_0, V_0, and ω_z we use (19) to obtain W_0. Finally, we obtain $\vec{N} = (p \ q \ 1)^T(1 + p^2 + q^2)^{-\frac{1}{2}}$ and $\|\vec{\omega}\| = \sqrt{\omega_z^2 + p^2\omega_z^2 + q^2\omega_z^2}$.

We have estimated the translational velocity $\vec{T}$ and the rotational velocity $\vec{\omega}$ in the camera coordinate system $Oxyz$. We are interested in the translational and rotational velocity expressed in the Frenet-Serret frame $Otnb$. By comparing equations (2), (8) and (9) we obtain

$$\vec{\omega} = v\kappa\vec{\mathbf{b}}, \quad \vec{\mathbf{b}} = \vec{N}\mathrm{sgn}\omega_z, \quad v\kappa = \|\vec{\omega}\| \tag{28}$$

where sgn stands for the 'sign of' function. Also, from (2), (8) and (9) we have

$$(U_0 \ V_0 \ W_0)^T = Z_0^{-1}\vec{T} = Z_0^{-1}(\dot{\vec{d_c}} - \vec{\omega} \times \vec{d_c}) = Z_0^{-1}(v\vec{t} - \vec{\omega} \times \vec{d_\gamma})$$

and thus

$$\frac{v\vec{t}}{Z_0} = (U_0 \ V_0 \ W_0)^T + \frac{\vec{\omega} \times \vec{d_\gamma}}{Z_0}. \tag{29}$$

Note that in equation (29) the quantities Z_0 and $\vec{d_\gamma}$ (the position of the point C, the origin of the $Otnb$ frame) are not known. However, let $\vec{d_\gamma} = (X_c \ Y_c \ Z_c)^T$ be the position of C and let (x_c, y_c) be the image of C (either the tip or the center of mass of the tool). From (18) we obtain

$$\frac{fZ_0}{Z_c} = f + px_c + qy_c$$

so that (29) can be written as

$$\frac{v\vec{t}}{Z_0} = (U_0 \ V_0 \ W_0)^T + \frac{Z_c}{fZ_0}\vec{\omega} \times \frac{f(X_c \ Y_c \ Z_c)^T}{Z_c} = (U_0 \ V_0 \ W_0)^T + \frac{\vec{\omega} \times (x_c \ y_c \ f)^T}{f + px_c + qy_c}. \tag{30}$$

9

From (30) we obtain the unit vector in the tangent direction $\vec{t}$ by normalizing $v\vec{t}/Z_0$. Finally, we obtain the unit vector in the normal direction using

$$\vec{n} = \vec{b} \times \vec{t}. \tag{31}$$

Equations (28), (30) and (31) define the Frenet-Serret frame $Otnb$ expressed in the camera coordinate system. Equation (28) gives us the curvature κ up to an unknown factor v (linear velocity). We conclude that the Frenet-Serret motion can be recovered up to the speed v; note that the translational velocity $v\vec{t}/Z_0$ does not help here because of the unknown depth Z_0.

Finally, we need to recover the orientation of the tool coordinate frame (its long and short axes) in the $Otnb$ frame. We find the long and the short axes of the tool as the principal axes of the set of tool points. The long axis l of the tool and the origin O of the fixed (camera) coordinate frame $Oxyz$ define a plane Π_l. Since the image l' of l lies in this plane we can find Pi_l using l' in place of l. Because we have assumed a nondegenerate view we have two cases: (i) if the tangent vector $\vec{t}$ lies in Π_l the motion is along l; (ii) if the normal vector $\vec{n}$ lies in Π_l the motion is orthogonal to l.

We check if the vector lies in the plane Π_l using the following simple algorithm. Let $\vec{p}_1 = (x_1 \ y_1 \ f)^T$ and $\vec{p}_2 = (x_2 \ y_2 \ f)^T$ be the positions of two endpoints on the image l' of l. The normal $\vec{N}_\Pi$ of the plane Π_l is given by

$$\vec{N}_\Pi = \vec{p}_1 \times \vec{p}_2.$$

If the vector $\vec{t}$ lies in the plane Π_l we have $\vec{N}_\Pi \times \vec{t} \approx 0$. So to find out the relative orientation of the tool frame and the $Otnb$ frame we only need to find which one of the inner products $|\vec{N}_\Pi \cdot \vec{t}|$ and $|\vec{N}_\Pi \cdot \vec{n}|$ is smaller. (Note that while one of the vectors $\vec{t}$ and $\vec{n}$ lies in the plane Π_l the other vector is not always orthogonal to Π_l .)

6 Vehicle Motion

We assume that the motion of the vehicle is planar and that it has a small rotational velocity around the axis orthogonal to the plane of motion. The translational velocity is dominant and at any time t the motion can be approximated by pure translational motion.

6.1 The Image Motion Field of a Moving Vehicle

From (14–15) we obtain the (approximate) equations of projected motion for points on a vehicle under weak perspective:

$$\dot{x} = \frac{Uf - xW}{Z_c}, \tag{32}$$

$$\dot{y} = \frac{Vf - yW}{Z_c}. \tag{33}$$

10

Equations (32-33) relate the image (projected) motion field to the scaled translational velocity $Z_c^{-1}\vec{T} = Z_c^{-1}(U\ V\ W)^T$.

Given the point $\vec{r} = x\vec{i} + y\vec{j}$ and the normal direction $n_x\vec{i} + n_y\vec{j}$, from (32-33) the normal motion field $\dot{\vec{r}}_n \cdot \vec{n} = n_x\dot{x} + n_y\dot{y}$ is given by

$$\dot{\vec{r}}_n \cdot \vec{n} = n_x f U Z_c^{-1} + n_y f V Z_c^{-1} - (n_x x + n_y y)W Z_c^{-1} \tag{34}$$

Let

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \equiv \begin{pmatrix} n_x f \\ n_y f \\ -n_x x - n_y y \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \equiv \begin{pmatrix} U Z_c^{-1} \\ V Z_c^{-1} \\ W Z_c^{-1} \end{pmatrix}. \tag{35}$$

Using (35) we can write (34) as $\dot{\vec{r}}_n \cdot \vec{n} = \mathbf{a}^T\mathbf{c}$. The column vector $\mathbf{a}$ is formed of observable quantities only, while each element of the column vector $\mathbf{c}$ contains quantities which are not directly observable from the images. To estimate $\mathbf{c}$ we need estimates of $\dot{\vec{r}}_n \cdot \vec{n}$ at three or more image points.

6.2   Estimating Vehicle Motion from Normal Flow

As in Section 5.2 we use linear least squares to estimate parameter vector $\mathbf{c}$ from the normal flow.

In the case of a moving vehicle the parameters of interest are the vehicle's trajectory and its *rate of approach*. The rate of approach

$$\nu = \frac{W}{Z_c}$$

(measured in sec^{-1}) is equivalent to the inverse of the *time to collision* and corresponds to the rate with which an object is approaching the camera (or receding from it). The rate $\nu = 0.1/\text{sec}$ means that every second the object travels 0.1 of the distance between the observer and its current position. A negative rate of approach means that the object is going away from the camera.

The direction of motion $\mathbf{c} \equiv \vec{T}/Z_c$ gives us the tangent vector $\vec{t} = \mathbf{c}/\|\mathbf{c}\|$ of the Frenet-Serret frame. If the direction of motion changes over time we can use the Frenet-Serret formulas (5) to recover the (scaled) curvature $\nu\kappa$ of the trajectory. Given the tangent direction $\vec{t}_0$ at time t and the tangent direction $\vec{t}_1$ at time $t + \Delta t$ we have

$$\vec{n}_0 = \nu\kappa\vec{n} \approx \frac{\vec{t}_1 - \vec{t}_0}{\Delta t}. \tag{36}$$

The unit vector in the direction $\vec{n}_0$ at time t is the normal vector of the $Otnb$ frame and the scaled curvature is given by $\nu\kappa = \|\vec{n}_0\|$. Finally, we obtain

$$\vec{b} = \vec{t} \times \vec{n}. \tag{37}$$

Equations (36) and (37) give us the normal $\vec{b}$ to the plane of motion and the rotational velocity of turning (yaw) $\vec{\omega} = \nu\kappa\vec{b}$.
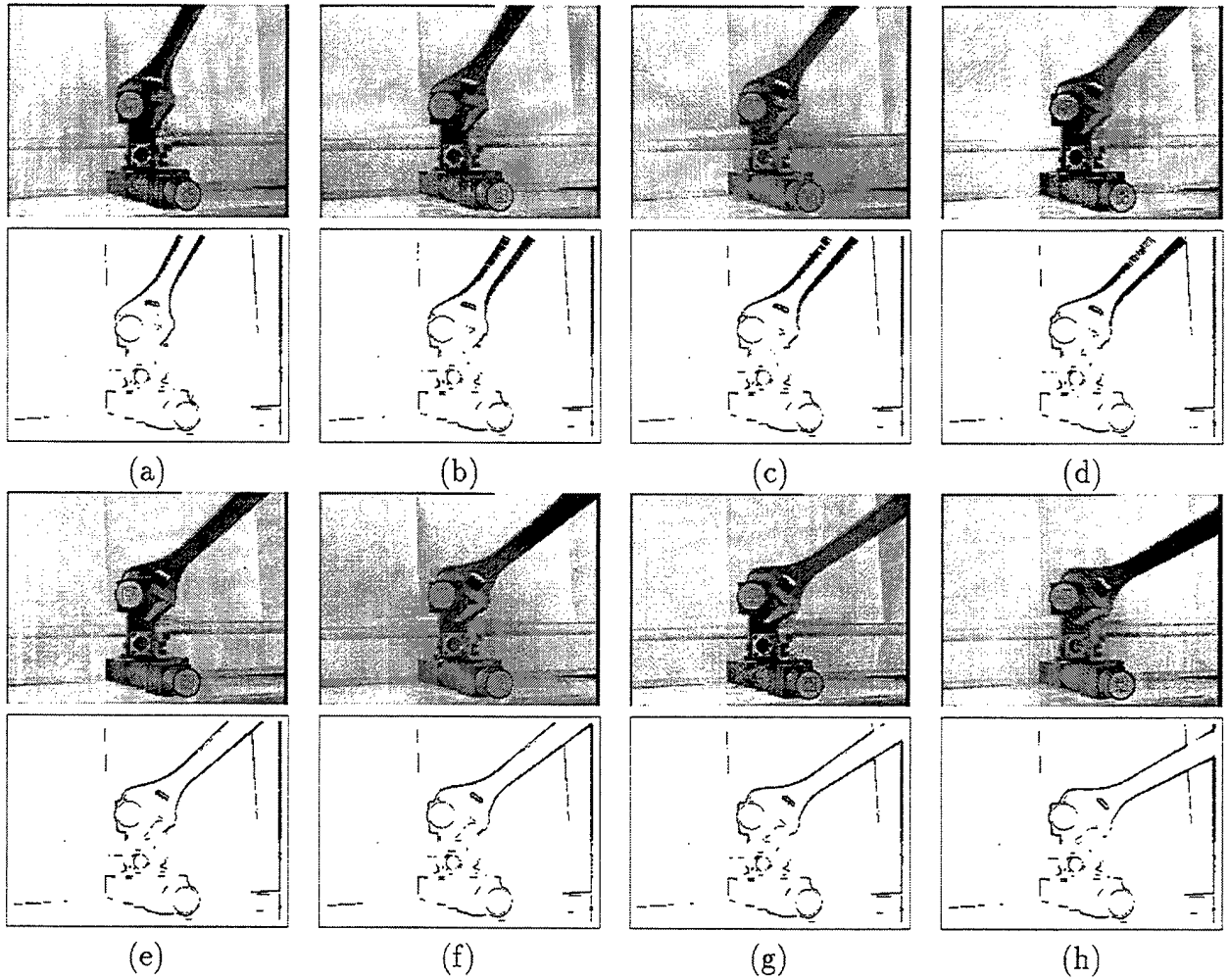
Figure 3: An experiment using a wrench: (a-h) frames 30, 40, ..., 100. Top images: the input images. Bottom images: results of flow computation.

7 Experiments

In the following section we show two examples for each of the domains we have discussed: tools and vehicles. As was mentioned before, tools usually operate by planar motion, advancing along a line (drill) or moving in a plane (sawing). In our examples we show two types of motion: rotation with negligible translation, and relatively small rotation and dominant translation. In Section 7.1 we will analyze saw and wrench examples.

A ground vehicle's motion usually takes place on terrain that has a small slope and on a road with a limited rate of turn. This results in small values of pitch and yaw, i.e. in locally planar, translational motion. Long sequences are needed to detect basic maneuvers such as turning or lane changing. In Section 7.2 we analyze two examples: an accelerating van (essentially linear motion) and a turning taxi.
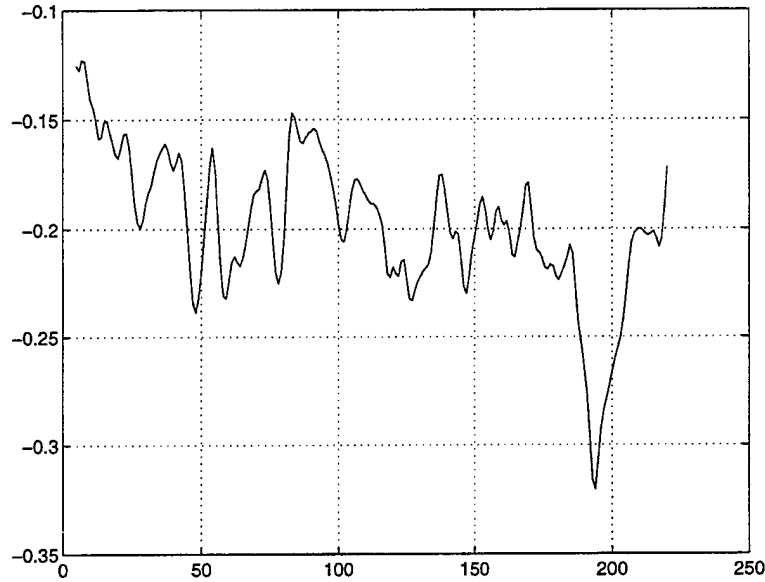
12

Figure 4: Results of experiments on the wrench sequence: the graph shows rotational velocity in radians/sec.

7.1 Motions of Tools

We tested our motion analysis algorithm under full perspective on two image sequences of tools in motion. The first sequence, shown in Figure 3, was a 200-image sequence of the movement of a wrench tightening a bolt.

The motion of the wrench was a rotation (to turn the bolt) around an axis approximately orthogonal to the plane of the image. The rotational velocity is shown in Figure 4; it is given in radians/sec and it corresponds to the scaled curvature $v\kappa$. Figure 5 shows the orientation of the principal axis of the wrench and the instantaneous translational velocity vector of its centroid (obtained using equation (30)), both measured in radians. As we see, the translational velocity vector remains approximately orthogonal to the principal axis throughout the motion sequence. The Frenet-Serret frame has its binormal $\vec{b}$ in the direction of the negative z-axis, its tangent $\vec{t}$ in the image plane and orthogonal to the principal axis of the wrench, and its normal $\vec{n}$ in the image plane and oriented from the centroid of the wrench toward the bolt.

We also tested our motion analysis algorithm on a 200-image sequence of a saw doing a periodic motion. Figure 6 presents part of the sequence. Flow results are given below each image. The motion of the saw was pure translation ($\|\vec{\omega}\| = 0$). As can be seen from Figure 7 the motion is mostly fronto-parallel (the z component of the translational velocity is small). The motion is periodic in the direction of the principal axis of inertia. It is a simple case of a (periodic) straight-line motion with the Frenet-Serret frame corresponding to the principal axes of the saw; $\vec{t}$ corresponds to the longest axis, and $\vec{b}$ to the shortest axis.

These graphs show that the motion components have a simple behavior; before they reach their extremal values they can be approximated by straight lines, indicating constant relative
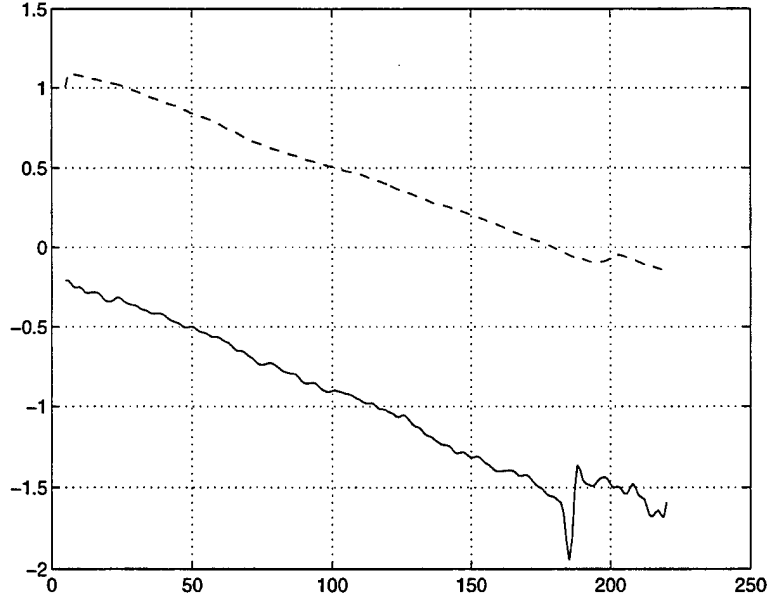
Figure 5: Results of experiments on the wrench sequence. The solid line corresponds to the orientation (in radians) of the instantaneous direction of translation of the centroid of the wrench, and the dashed line corresponds to the orientation (in radians) of the principal axis of the wrench.

accelerations.

7.2 Motions of Vehicles

For vehicle motion we also used two image sequences, and we used the algorithms for weak perspective. In the first experiment we used an image sequence of a van taken from another vehicle following the van. The sequence consisted of 56 frames (slightly less than two seconds). Figure 9 shows frames 5, 15, 25, and 35 as well as the corresponding normal flow on the van. Figure 10 shows estimated values of UZ_c^{-1}, VZ_c^{-1}, and WZ_c^{-1}. These values correspond to the relative translation of the van and the vehicle carrying the camera (observer coordinate system). Because of our choice of the coordinate system the rate of approach ν corresponds to the negative of WZ_c^{-1}, i.e. $\nu = -WZ_c^{-1}$. The graph shows that there is an impending collision (rate of approach greater than 1 sec^{-1}). Around the 20th frame the rate of approach becomes zero (as do all the velocity components) and after that it becomes negative because the van starts pulling away from the vehicle carrying the camera.

In the second experiment we used an image sequence of a turning taxi taken by a stationary camera. The sequence consisted of 21 frames. Figure 11 shows frames 1,9,15 and 21 as well as the corresponding normal flow on the vehicles. Figure 12 shows estimated values of UZ_c^{-1}, VZ_c^{-1}, and WZ_c^{-1}. These values correspond to the relative translation of the taxi. The graph shows that there is a large W component in the turn (the taxi is receding), and that the turn is to the right (negative U, positive V).
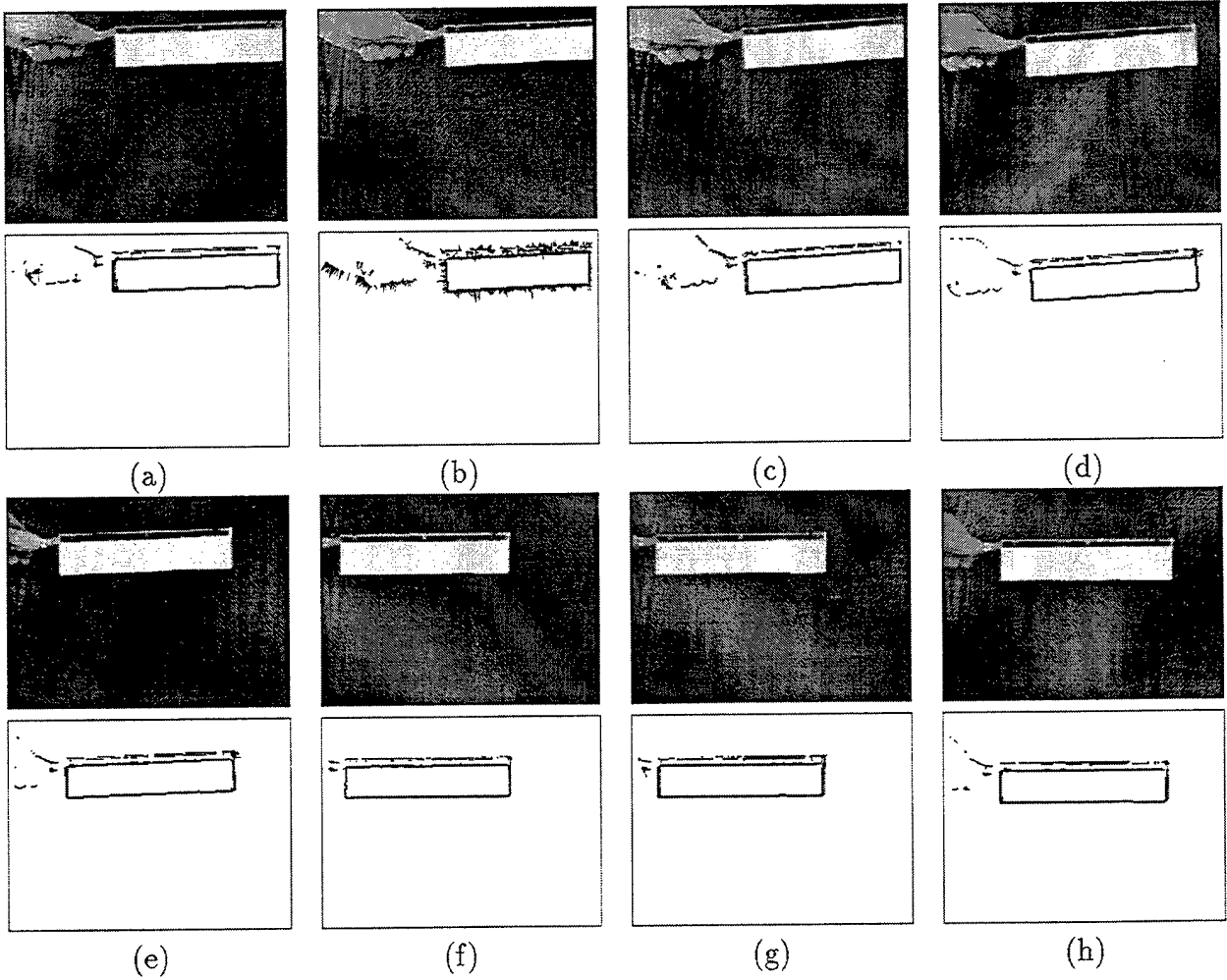
14

Figure 6: An experiment using a saw: (a-h) frames 30, 40, ..., 100. Top images: the input images. Bottom images: results of flow computation.

8 Conclusions

Many types of common objects, such as tools and vehicles, usually move in simple ways when they are wielded or driven: The natural axes of the object tend to remain aligned with the local trihedron defined by the object's trajectory. In this paper we have considered the relationship between this constrained motion and the object's geometry. To analyze this relationship we have used two frames: the object frame and the frame of the motion trajectory. Assuming a constant relationship between the object frame and the motion frame during the motion, we have used *Frenet-Serret motion* as a motion model. Using the Frenet-Serret frame has provided us with an ability to understand motion over a long time period.

We have derived equations for understanding the motions of tools and vehicles under full and weak perspective. We have recovered descriptions of an object's motion and the space
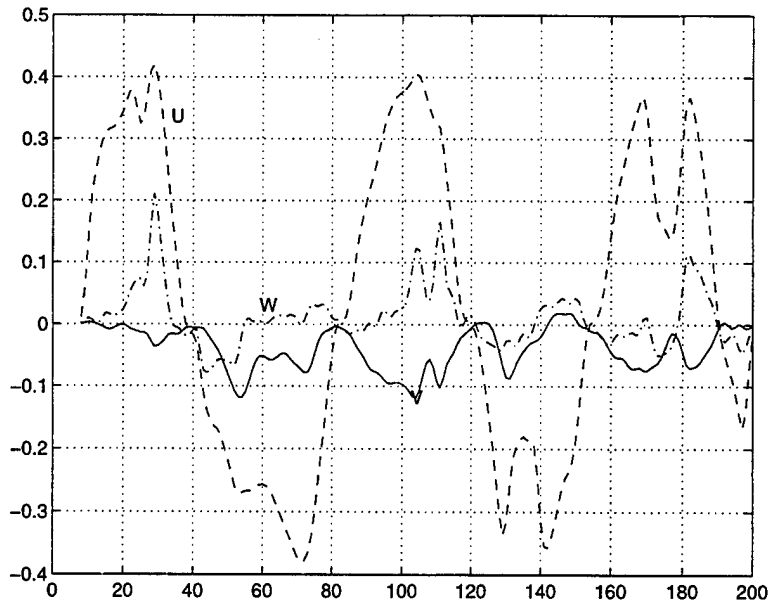
Figure 7: Results of experiments on the saw sequence. U, V, W are the scaled (by an unknown distance Z_0^{-1}) components of the relative translational velocity.

curve along which the object moves, using relatively long image sequences. The motion and trajectory parameters provide a low-level description for understanding the motions of vehicles. For understanding tools in motion one needs additional knowledge about the tool and the context. This is a direction for further research.

It is the need for efficient force transfer that imposes a simple and constant relationship between the natural axes of the object and the motion trajectory. We have used this functional constraint in analyzing the motions of tools and ground vehicles. Expanding this analysis to other classes of objects (e.g. air vehicles), as well as expanding the vocabulary that describes the behavior of tools and vehicles (sharp turn, skid, etc.) [13] are other directions for future research.

References

[1] T. J. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:90–99, 1986.

[2] A. M. Bruckstein, R. J. Holt, and A. N. Netravali. How to catch a crook. *Journal of Visual Communication and Image Representation*, 5:273–281, 1994.

[3] A. M. Bruckstein, R. J. Holt, and A. N. Netravali. How to track a flying saucer. *Journal of Visual Communication and Image Representation*, 7:196–204, 1996.

[4] H. Buxton and R. Howarth. Watching behaviour: The role of context and learning. In *Proc. International Conference on Image Processing*, volume 2, pages 797–800, 1996.
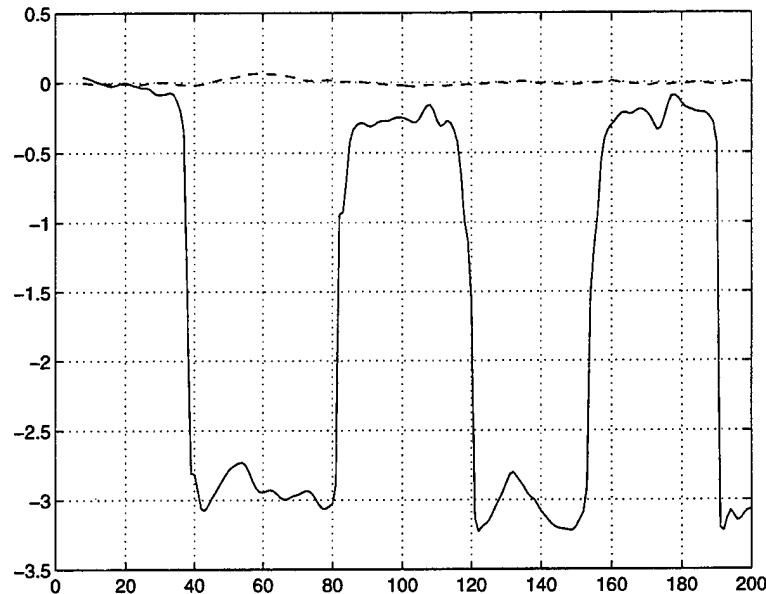
Figure 8: Results of experiments on the saw sequence. The solid line corresponds to the orientation (in radians) of the instantaneous direction of motion of the saw, and the dashed line corresponds to the orientation (in radians) of the principal axis of the saw.

[5] C. Cedras and M. Shah. Motion-based recognition: A survey. *Image and Vision Computing*, 13:129–155, 1995.

[6] Z. Duric, J. Fayman, and E. Rivlin. Function from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:105–122, 1995.

[7] Z. Duric and A. Rosenfeld. Image sequence stabilization in real time. *Real-Time Imaging*, 2:271–284, 1996.

[8] Z. Duric, A. Rosenfeld, and L. S. Davis. Egomotion analysis based on the Frenet-Serret motion model. *International Journal of Computer Vision*, 15:105–122, 1995.

[9] S. Engel and J. Rubin. Detecting visual motion boundaries. In *Proc. Workshop on Motion*, pages 107–111, 1986.

[10] G.H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, 10:413–432, 1973.

[11] K. Gould and M. Shah. The trajectory primal sketch: A multi-scale scheme for representing motion characteristics. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 79–85, 1989.

[12] N. Johnson and D. Hogg. Learning the distribution of object trajectories for event recognition. *Image and Vision Computing*, 14:609–615, 1996.
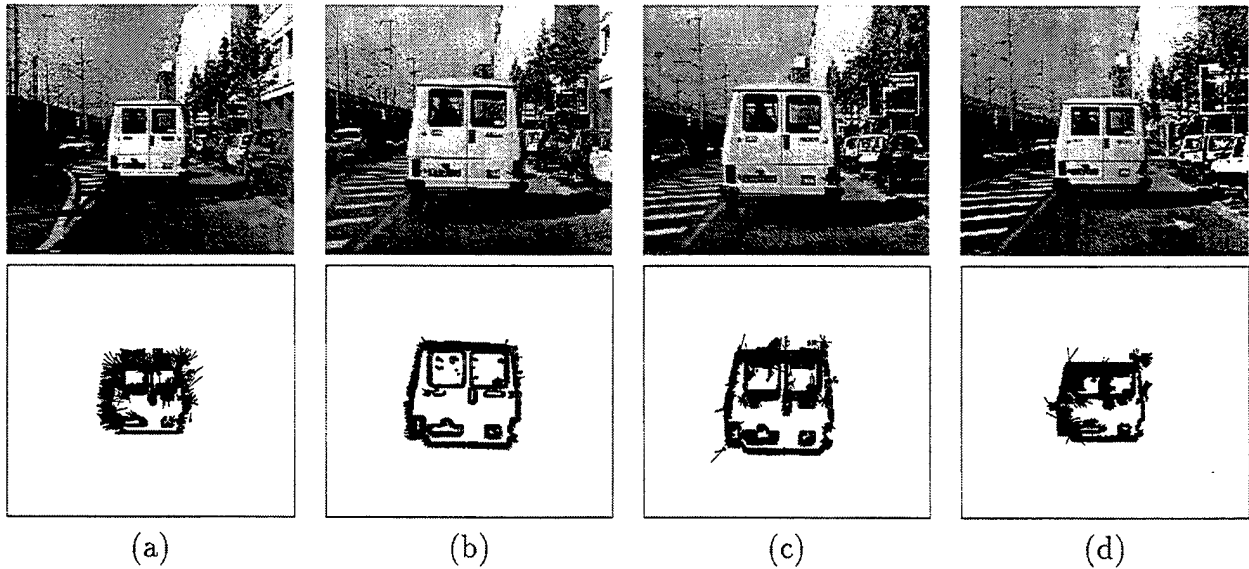
Figure 9: Frames 5, 15, 25, and 35 of the van sequence. The normal flow results are shown below the corresponding image frames.

[13] D. Koller, H. Heinze, and H.H. Nagel. Algorithmic characterization of vehicle trajectories from image sequences by motion verbs. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 90–95, 1991.

[14] E. Kreyszig. *Differential Geometry*. University of Toronto Press, Toronto, Canada, 1959.

[15] A. Verri and T. Poggio. Against quantitative optical flow. In *Proc. International Conference on Computer Vision*, pages 171–180, 1987.

[16] J. Weng, T.S. Huang, and N. Ahuja. 3-D motion estimation, understanding, and prediction from noisy image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:370–389, 1987.
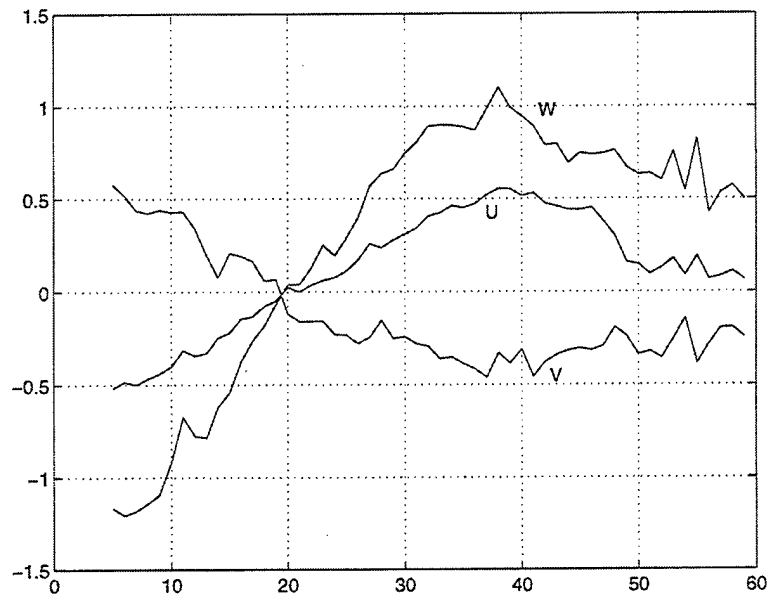
Figure 10: Results of experiments on the van sequence. U, V, W are the scaled (by an unknown distance Z_c^{-1}) components of the relative translational velocity .

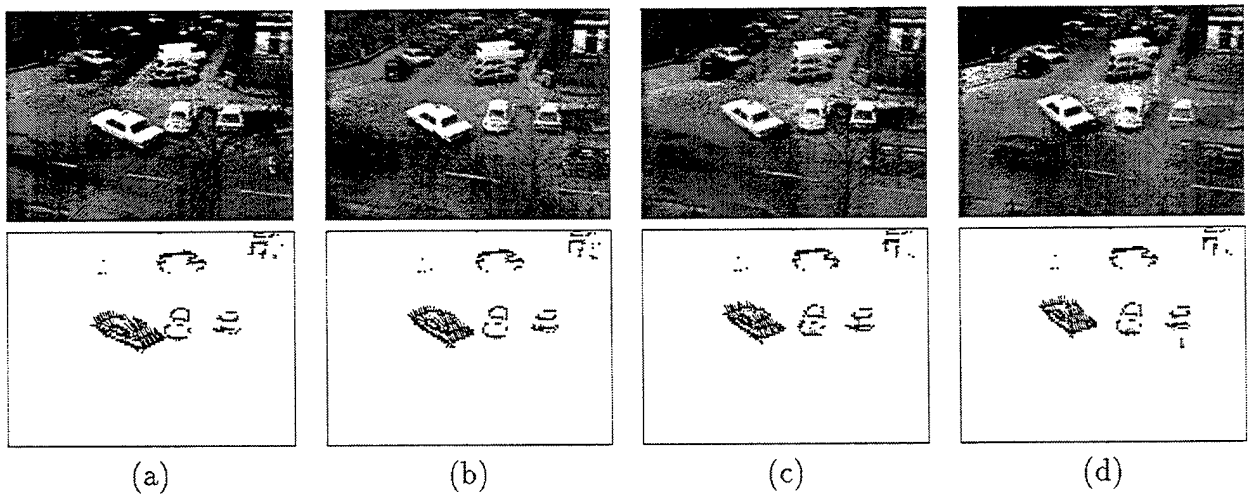|			
(a)	(b)	(c)	(d)

Figure 11: A taxi sequence: (a-d) frames 1, 9, 15, 21. Top images: the input images. Bottom images: results of flow computation.
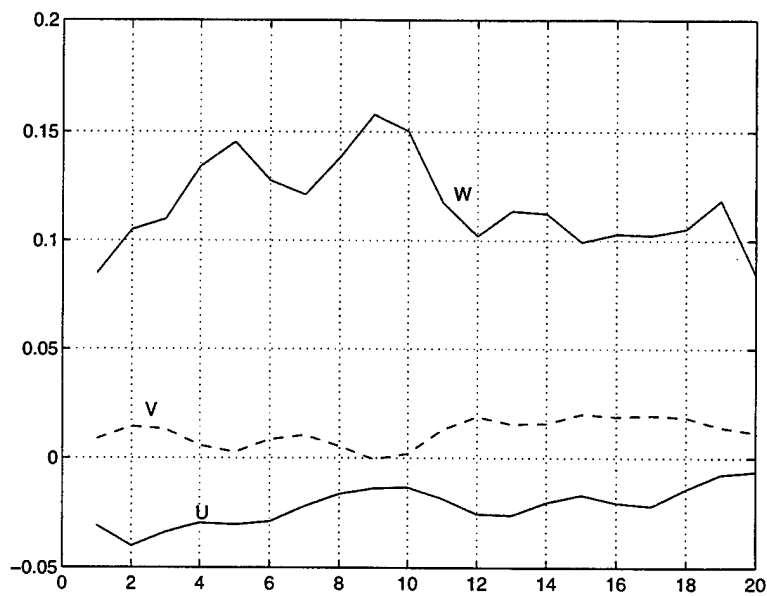
Figure 12: Results of experiments on the taxi sequence. U, V, W are the scaled (by an unknown distance Z_c^{-1}) components of the relative translational velocity .

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE March 1997	3. REPORT TYPE AND DATES COVERED Technical Report

4. TITLE AND SUBTITLE

Understanding Object Motion

5. FUNDING NUMBERS

N00014-95-1-0521

6. AUTHOR(S)

Zoran Duric, Ehud Rivlin, and Azriel Rosenfeld

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Center for Automation Research
University of Maryland
College Park, MD 20742-3275

8. PERFORMING ORGANIZATION REPORT NUMBER

CAR-TR-854
CS-TR-3780

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street, Arlington, VA 22217-5660

Advanced Research Projects Agency
3701 North Fairfax Drive, Arlington, VA 22203-1714

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release.
Distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Many types of common objects, such as tools and vehicles, usually move in simple ways when they are wielded or driven: The natural axes of the object tend to remain aligned with the local trihedron defined by the object's trajectory. Based on this observation we use a model called *Frenet-Serret motion* which corresponds to the motion of a moving trihedron along a space curve. Knowing how the Frenet-Serret frame is changing relative to the observer gives us essential information for understanding the object's motion. This is illustrated here for four examples, involving tools (a wrench and a saw) and vehicles (an accelerating van, a turning taxi).

14. SUBJECT TERMS

Frenet-Serret motion, Normal optic flow, Nonlinear least squares, Motion models, Object dynamics, Object motion, Vehicle motion, Tool motion, Weak perspective projection, Planar motion

15. NUMBER OF PAGES
23

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to *stay within the lines* to meet *optical scanning requirements*.

Block 1. Agency Use Only (Leave blank).

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	-	Contract	PR	- Project
G	-	Grant	TA	- Task
PE	-	Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. (If known)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA - Leave blank.
NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17.-19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.